LDWSF 2.8.1
07/17/06

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
REGION 10
1200 Sixth Avenue
Seattle, WA 98101

July 17, 2006

Reply to
Attn Of: ECL-111

Mike Johns
Windward Environmental
200 W. Mercer Street, Suite 401
Seattle, WA 98119

Re: Comments on Food Web Model Memorandum 3: Preliminary Model Results; Lower Duwamish Waterway Superfund site; Seattle, Washington

Dear Mike:

Enclosed are comments from the Duwamish River Cleanup Coalition (DRCC) on the Lower Duwamish Waterway Group's Food Web Model Memorandum 3: Preliminary Model Results, dated April 7, 2006. EPA and Ecology's comments included consideration of DRCC's comments. DRCC's specific comments should be considered as we move forward with a technical workgroup to resolve remaining issues related to the FWM.

Please contact me at (206) 553-2140 or Bruce Duncan at (206) 553-0218 if you have any questions.

Sincerely,

Allison Hiltner
Superfund Site Manager

cc: Brad Helland, Ecology (hard copy)
(electronic copies only):
Jennie Goldberg, City of Seattle
Jeff Stern, King County
Skip Fox, Boeing Company
Doug Hotchkiss, Port of Seattle

Duwamish River Cleanup Coalition

Community Coalition for Environmental Justice • The Duwamish Tribe • ECOSS • Georgetown Community Council • IM-A-PAL • People for Puget Sound • Puget Soundkeeper Alliance • South Park Neighborhood Association • Washington Toxics Coalition • Waste Action Project

May 4, 2006
Allison Hiltner
hiltner.allison@epamail.epa.gov

Dear Allison:

The Duwamish River Cleanup Coalition has reviewed the Lower Duwamish Waterway Group's Draft Food Web Model Deliverable #3. Our review indicates that the application and development of the model appear reasonable, but raises several questions and comments regarding the accuracy assessment and sensitivity analysis of the model.

1. The report states that the analysis was conducted to decide whether additional water column data should be collected this summer. What is the (tentative) conclusion to this?

2. The concentrations of PCBs in sediment are being interpolated and this is ongoing. The comparisons between model results and empirical data are likely to be different with any modification to the interpolated data set. Will model performance criteria be evaluated and Monte Carlo simulation performed again when input parameters have been fully developed?

3. The report states that 20 parameters values were defaults values. How sensitive is the model to these parameters? How reasonable are the assumptions?

4. In what way are PCBs different between site-wide and smaller area modeling applications? The term "scale" in the report refers to the geographic scale, not map scale (it would be more accurate to refer to a smaller "area," not smaller "scale.") However, the accuracy assessment and sensitivity analysis are somewhat misplaced given that the resolution of the data does not increase with decreasing area.

5. What is the reason for having two aspatial model performance criteria (SPAF and % difference)? These metrics say the same thing, but SPAF is potentially confusing. To be conservative, the model should over-predict (within the accepted accuracy threshold). So why not just use SPAF = PTCC/ETCC? This would convey at a glance what is over-predicted (values greater than 1) and under-predicted (under 1). Otherwise, for future reporting we suggest the analysis only report % difference.

6. Table 3-3 illustrates how misleading the unsigned SPAF metric is. It appears that the 2004 and 2005 data predictions for shiner surfperch and English sole are similar (1.3 to 1.8; 1.1 to 1.7), but in there is a large (~80–100%) shift in comparing against one data set and the other, illustrated by the % difference metric (-24% to 81%; -11% to 72%).

7. The model performance metrics measure relative difference, but not absolute difference. For example, if an empirical value for a species is 30, with 60 being predicted, the relative difference is 100%, but the absolute difference is 30. If the empirical value is 3 and predicted is 6 the relative difference is also 100%, but the absolute difference is only 3. Is the absolute difference significant for decision making? Is there some critical absolute threshold of tissue concentration that is significant?

8. SPAF is computed using mean values. Thus, in some cases the model prediction is much further off the mean than is apparent in the SPAF value. We recommend considering use of median values, so that (a) outliers have a smaller influences and (b) SPAF says something about the number sample values that are close to the comparison value (i.e., the median).

9. The stated goal for SPAF is 3, with a requirement of 5. Within the empirical data there are max values that are more than 10 times greater than the minimum value. Why would you a) expect the model to do better that the empirical variance and b) think that the empirical data can provide proof of accuracy greater than it's internal validity?

10. Neither of the model performance criteria account for either point or spatial variability. A better comparison would relate empirical data distributions with modeled (Monte Carlo) distributions, either aspatial or spatially or both.

11. Where is it the comparison of different interpolation methods for PCB concentrations reported?

12. Whether it is appropriate to use the extreme value for shiner surfperch in the calculation of the mean should be considered, especially in comparison to the historical and 2005 data. Is it possible that it was a measurement error? If so, there may be little evidence to support spatial averaging of this value. That is, is the value representative of the sample location and is it appropriate to apply such an extreme value for the whole site? The most accurate prediction listed in Table 3-2 is for the shiner surfperch, which is associated with the most extreme sampled value. Obviously, if the extreme value is dropped, the modeled accuracy will change. Windward seems to suggest that the value should be dropped:

> "The range of empirical total PCB concentrations for shiner surfperch was much greater than that for other species, and the model predictions did not bound this range (Table 6-1). In particular, there was one shiner surfperch sample with an exceptionally high concentration (18,000 g/kg). The Monte Carlo model predictions did not bound the highest empirical total PCB concentration for shiner surfperch." (p. 43)

13. Table 3-2 shows that predictions for benthic invertebrates are accurate within a factor of 1.8, however the "empirical data" is simulated using a regression equation. The regression equation is log-log and only has an r-squared value of 0.74, which is not a high value (though not surprising). Thus it must be recognized that the FWM predicts values on average within 1.8 times of another model that predicts empirical data with considerably less accuracy. This illustrates the importance of the sensitivity analysis and agreement on model input data and

parameters, more so than significance or agreement on an appropriate SPAF.

14. Table 3-2 shows that predictions for slender crab are accurate within a factor of 1.4. However, the empirical data set includes no historical data and has the fewest total samples (across 2004 and 2005) of any species. It must be understood and explained that there are inconsistencies in the data sets between species. The reason for this should be explained.

15. For Figure 3-1, instead of the 1st/3rd quartile bar, we recommend showing the +/- 3x or 5x range for the SPAF to illustrate (a) the magnitude of the this range and (b) it's relative variation with increasing absolute values.

16. Table 3-3 (separate comparison to 2004 and 2005 data) shows large differences between data sets for species (crab) with small sample sizes, but not for the species with larger sample sizes. This suggests that the number of samples for crab is not capturing population variability.

17. Table 4-2, which shows model results across different diet scenarios, is not reasonably interpreted. The value (benefit) in analyzing the scenarios is not a determination of the most accurate scenario, but the sensitivity of the model to dietary scenarios. For three of the species, it clearly doesn't matter, within the accuracy of the model, what scenario is used. This alleviates the need to focus a lot of effort on deciding upon the dietary scenarios for these species. For crab, sole, and particularly sculpin, the model is sensitive to different dietary inputs. Further thought, discussion and/or data collection should be done for these species.

18. For the sensitivity analysis, Arnot (2006) is cited to say that the model is considered "highly sensitive" when the input value change to model output change is great than 1:1 (i.e., a 0% difference), but then 8% is selected as the threshold for inclusion into sensitivity analysis for the 10% difference analysis and 10% for the plausible range analysis. Why is 8% used for the uncertainty analysis selection criteria for the 10% sensitivity analysis and 10% for the plausible range analysis? There is no apparent reason for this; the analyses are basically the same thing, with the latter looking at a larger % difference.

19. Why is pacific staghorn sculpin so much more sensitive to dietary absorption efficiency of lipids than any other species (alpha) in the 10% sensitivity analysis? Why is slender crab so much more sensitive to water content than the other species? (For all other parameters, the max, min, and mean are quite similar).

20. Table 5-2, reporting the plausible range sensitivity analysis, is impossible to interpret. The SPD figures are not comparable (as noted in the report) because the percent changes applied to the parameter are different across species and across parameters. Only Table 5-3 is necessary. The relative response ratio should be reported as a percent, like the 10% sensitivity analysis results for easy comparison. For clarity, we recommend eliminating Table 5-2.

21. It's troubling that some parameters and input data were left out of the sensitivity analysis. The justification is that using the software that Windward has, the Monte Carlo simulation can't be performed on equation-based values. But this does not mean that a sensitivity analysis can't be performed and isn't necessary. The value (benefit) of a sensitivity analysis is insight into the

model's behavior with respect to the various inputs and parameters to inform decisions about data collection and fitness of data sets for modeling (i.e., if the model is insensitive to the input, then a large spread in empirical input data is not of great concern). The uncertainty analysis is not the primary reason for conducting the sensitivity analysis.

22. In fact, the uncertainty analysis does not provide the majority of insight claimed in the report; rather, the sensitivity analysis does. Monte Carlo does not reveal attributes of a model; it is a means of propagating uncertainty in the inputs to model outputs. The demonstration of Monte Carlo simulation is encouraging. Representing results as a cumulative distribution function is most useful. In the future, SQTs etc. can be chosen based on associated exceedance probabilities.

23. The best indicator of the model's performance in the report is not SPAF or % difference, but the comparison of the Monte Carlo results with the empirical data sets. In fact, rather than SPAF, a more formal statistical comparison of modeled distributions to empirical distributions should be done (e.g., ANOVA). At the very least, more emphasis should be put on the analysis as described on p. 43:

> "In addition, the 5th and 95th percentiles of predicted tissue concentrations were, with the exception of shiner surfperch, within a factor of 3 or better of the empirical minimum and maximum concentrations, respectively."

24. Applying the model to the smaller areas is not necessarily an evaluation of the model. It's more an evaluation of data resolution and spatial variability of either the empirical data or input data. The results suggest that there is significant spatial variability, but Windward does not make an effort to determine if this is an attribute of the empirical tissue data, model input data or both. More than likely it's related to the empirical data set of tissue concentrations because of the high sample sizes for input data. There is no evidence cited to suggest that the average values over the larger area are equal to supporting values for the modeling areas. It further seems to suggest that the resolution of data does not support modeling at the smaller geographic scales. Averaged over the entire LDW-site, there are obviously compensating "errors" (variability), leading to higher accuracy values. Because the model performance criteria is not a spatial criterion, application of the model to the smaller modeling areas is the only insight provided of fidelity of modeled spatial variance. The FWM model is not a spatial model (there is no spatial dependency). Thus if spatial patterns are important, either the empirical data or input data resolution does not appear to be high enough.

Thank you for the opportunity to review the Draft Food Web Model Technical Memo #3. We look forward to reviewing the final report.

Sincerely,

BJ Cummings

BJ Cummings